

# **FindDefault (Prediction of Credit Card fraud)**

## **Problem Statement:**

A credit card is one of the most used financial products to make online purchases and payments. Though credit cards can be a convenient way to manage finances, they also come with risks such as credit card fraud. This is the unauthorized use of someone else's credit card or credit card information to make purchases or withdraw cash. The dataset contains transactions from September 2013 by European cardholders over two days with 492 frauds out of 284,807 transactions.

This dataset is highly unbalanced as the frauds account for only 0.172% of all transactions. A classification model is needed to predict fraudulent transactions to safeguard customers from illegitimate charges.

## **Methodology:**

The predictive model was developed following these steps:

1. Data Preprocessing: Handled duplicates, missing values, and outliers.
2. Feature Engineering: Selection and scaling of relevant features.
3. Data Balancing: Applied techniques to address the imbalance in the dataset.
4. Train/Test Split: Segregated data into a training set and a testing set.

## **Model Training:**

The Random Forest classifier was chosen due to its robustness and the ability to handle unbalanced datasets. It's an ensemble learning method that operates by constructing multiple decision trees and outputting the class that is the mode of the classes of individual trees. It generalizes well and reduces the risk of overfitting.

## **Performance Evaluation:**

The model's performance was evaluated using various metrics including ROC AUC, precision, recall, and F1-score. These metrics were chosen to provide a holistic understanding of the model's ability to predict fraudulent transactions and minimize false negatives where a fraudulent transaction is misleadingly labelled as non-fraudulent.

## **Future Work:**

Future work could involve experimenting with other classification models, incorporating additional features that can potentially improve prediction accuracy, performing feature importance analysis, and developing a real-time detection system. Additionally, continuous model monitoring, updating, and re-training with new transaction data would be essential to maintain high performance.